

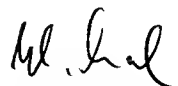
VERIFICATION OF TRANSLATION

I, Helmut Madl,
translator to

Winter, Brandl, Fürniss, Hübner,
Röss, Kaiser, Polte ■ Partnerschaft
Patent Attorneys and Attorneys at Law
of 85354 Freising, Germany,

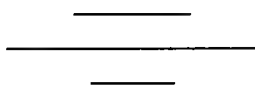
do solemnly and sincerely declare that I am conversant with the
English and German languages and am a competent translator
thereof, and that to the best of my knowledge and belief the
attached is a true and correct translation of a German Patent and
Trademark Office Priority Certification
in respect of **German Patent Application No. 10 2004 010 928.1**.

September 10, 2010



.....
(Helmut Madl)

FEDERAL REPUBLIC OF GERMANY



5

**Priority certificate on filing
a patent application**

10 **Serial number:** 10 2004 010 928.1

Application date: March 5, 2004

Applicant/Proprietor: SIRS-Lab GmbH, 07745 Jena/DE

15

Title: *Protein zur Bindung prokaryonter DNA sowie Verfahren
zur Trennung und Anreicherung prokaryonter DNA*
[Protein for binding prokaryotic DNA and method of
separating and enriching prokaryotic DNA]

20

IPC: C 07 K, C 12 N

**The annexed sheets are a true and exact reproduction of the original papers of this
patent application.**

25

Munich, March 3, 2005

By authority of

30

**The President of the
German Patent and Trademark Office**

(signed)

Dzierzon

35

SIRS-Lab GmbH
Attorney's file: PAT 3696/029

March 5, 2004
H/18/kt

**Protein for binding prokaryotic DNA and method of separating and
enriching prokaryotic DNA**

The invention relates to a protein which binds non-methylated cytidine-phosphate-guanosine dinucleotides (CpG motifs) of DNA, a nucleic acid coding therefor, a method of using the protein of the invention for separating and/or enriching prokaryotic DNA or depleting this DNA from physiological fluids, as well as to a kit for carrying out said methods.

Infections caused by bacteria are one of the most frequent causes of inflammatory diseases. Early detection of the bacterial pathogens is crucial for the prognosis of the course of the disease as well as, in particular, for timely selection of suitable therapeutic measures.

For the detection of bacterial pathogens, use is primarily made even today of different culture-dependent methods. As a result of the disadvantages of these methods, increased efforts were undertaken to find alternatives, especially during the past decade, simultaneously with the rapid technological development in molecular biology. First reports on the use of culture-independent methods of detecting bacterial pathogens, based on the principle of the polymerase chain reaction (PCR), date back to the early 1990s. Thus, for instance, Miller and colleagues (Miller N J Clin Microbiol. 1994 (Feb;32(2):393-7) were able to show that culture-independent methods are superior to the classical techniques of cultivation and microscopy in the detection of *mycobacterium tuberculosis*. Recently, however, further molecular-biological methods based on the detection of pathogen-specific nucleic acids have gained importance (e.g., M. Grijalva et al. Heart 89 (2003) 263-268; Uyttendaele M et al. Lett Appl Microbiol. 2003;37(5):386-91; Saukkoriipi A et al. Mol Diagn. 2003 Mar;7(1):9-15; Tzanakaki G et al. FEMS Immunol Med Microbiol. 2003 Oct. 24;39(1):31-6).

Besides the high specificity of such molecular-biological methods, the reduced time expenditure is to be mentioned as a substantial advantage over conventional culture-dependent methods. Nevertheless, the sensitivity of direct detection of prokaryotic DNA from body fluids and not from pre-treated testing material as compared to culturing of

microorganisms has been much too low so far. At best, an amount of nucleic acids of bacteria sufficient for the direct detection of pathogens from testing material which is not pre-treated is achieved in the area of 16S-rRNA analysis by means of PCR of the 16S region on the bacterial chromosome and the subsequent sequence analysis of the PCR fragment, because in most cases several copies for the segment coding the 16S-rRNA are situated on the chromosome. The direct specific detection of pathogens by means of 16S-rRNA analysis requires that only one pathogen species is present in the sample to be examined. If there are different pathogen species in the sample, specific detection by way of sequencing of the 16S-rRNA region is not possible because the primers used are universal for most bacteria. Furthermore, the pathogens to be detected must be in the metabolic phase and express enough 16S-rRNA.

As a rule this can not be assumed to be the case, in particular in patients subject to a calculated antibiotic therapy. Moreover, an expression of certain pathogenicity factors of bacteria does not occur at all times despite the fact that the corresponding genes are present in the bacterial genome. As a result, false negative results are transmitted to the clinical physician. Selective antibiotic therapy may thus be initiated either not at all or much too late. In such cases, the physician has to rely on his knowledge gained by experience and on general guidelines (such as those of the Paul Ehrlich Foundation) and will therefore effect a much too general antibiotic treatment. The unspecific use of antibiotics bears a number of risks, not only for the individual patient (such as unnecessary side effects in the form of renal damage etc.), but also for the entire society (e.g., the development of additional antibiotic resistances such as MRSA (methicilline-resistant *Staphylococcus aureus*, etc.). Therefore, the detection of clinically meaningful pathogenicity factors and resistances of bacteria on the chromosomal level and on the plasmid level, i.e. ultimately on the DNA level, provides considerable advantages for the diagnosis of many infectious diseases but also of sepsis. This applies even more because a distinction between pathogenic and commensal bacteria can also be made on this level.

Most frequently, the detection of pathogen-specific nucleic acids is effected by nucleic acid amplification techniques (NAT), such as the amplification of the prokaryotic DNA by means of the polymerase chain reaction (PCR) or the ligase chain reaction (LCR), respectively. The high specificity and fast availability of the results is contrasted by the susceptibility to interference by contamination or by strongly reaction-inhibiting factors in clinical samples.

In a conventional PCR detection method, successful detection of pathogens in the blood requires at least 1 target DNA of the pathogen to be present in 10 µl of blood. This corresponds to approximately 100 targets in 1 ml of blood or 1,000 targets in 10 ml of blood, respectively.

5

Things are different with the blood culture for the detection of pathogens causing an infection. In this case, the lower detection limit is approximately 3-5 bacteria per 10 ml of blood.

10 This detection limit is presently not reached yet by PCR methods, not even by those which have their target sequence in the area of the 16S-rRNA region on the chromosome. Although several regions coding 16S-rRNA - in most cases 3 to 6 - are located on the bacterial chromosome, there is still the prerequisite of at least one molecule of the template DNA being present in the PCR reaction mixture. Improved
15 diagnostic safety is to be expected of PCR methods whose specific target sequences code for species-specific proteins, either in the chromosome or on plasmids of the microorganisms. The above remarks with respect to the detection limit also apply here. Especially under the influence of an ongoing antibiotic therapy, growth of the pathogens can be decelerated or limited considerably even if the antibiotic employed ultimately does
20 not take effect. This situation is often found especially in patients who are already under antibiotic treatment and in whom disease-causing bacteria can therefore not be grown from the blood cultures or from other samples (such as, for example, tracheal smears, broncho-alveolar lavages (BAL), etc.).

25 Due to insufficient sensitivity, the pathogen-specific detection of nucleic acids in the absence of an amplification step by direct detection of prokaryotic DNA (probe technique, FISH technique) is of diagnostic importance only at a sufficiently high germ count in the test material.

30 Apart from PCR-inhibiting ingredients in the test material, the essential set of problems in the detection of prokaryotic DNA for the identification of bacterial pathogens in body fluids consists mainly in the excess amounts of eukaryotic DNA in contrast with prokaryotic DNA. In this regard, competitive processes in DNA analysis as well as the low quantity of prokaryotic DNA can in particular be regarded as a hindrance to qualitative and
35 quantitative detection of pathogens.

The usual methods of DNA isolation enrich the total DNA of a body fluid such that the ratio of host DNA to microbial DNA may be between $1:10^{-6}$ and $1:10^{-8}$. This difference makes the difficulty in detecting microbial DNA in body fluids quite easy to understand.

It would thus be desirable to be able to separate the prokaryotic DNA from eukaryotic DNA, and in particular to also enrich it relative to the latter.

- 5 The present invention accordingly is based on the object of providing means and methods allowing to separate and/or enrich prokaryotic DNA from examination samples having a high proportion of eukaryotic DNA, in particular of patients with infections.

10 In accordance with the invention, this is achieved through a protein which binds non-methylated CpG motifs while having a 25% to 35% homology, in particular about 27.6% homology, with the wild type CPGB protein, wherein the binding site for non-methylated CpG motifs is contained in the protein of the invention, and it is shortened in comparison with the wild type protein, preferably to the length of the binding site for non-methylated CpG motifs at maximum. In other words, the maximum shortening is only to such an
15 extent that the binding site for non-methylated CpG motifs is preserved.

What is referred to in the following as the wild type CPGB protein (or CPGbP656) is the human CPGB protein (cf. Voo et al., Mol Cell Biol. March 2000; 20(6): 2108-21). The protein of the invention shall in the following be referred to as CPGbP181. The protein
20 described in EP 02020904, which is a shortened variant of the wild type CPGB protein and served as the basis for the protein of the invention, shall in the following be referred to as as CPGbP241.

25 In the following, the invention will be described by making reference to the figures, wherein:

- Fig. 1 shows the amino acid sequence of CPGbP181 (**bold**) compared with the wild type CPGB protein (CPGbP656) and CPGbP241 (*italics*);
- 30 Fig. 2 shows the DNA sequence and translation to the amino acid sequence of the complete CPG-binding protein CPGbP656, wherein the shortened CPG-binding peptides CPGbP241 (**bold**) and CPGbP181 (*italics*) are represented;
- Fig. 3 shows a PCR of streptococci DNA in human blood;
- 35 Fig. 4 shows a nested PCR with the PCR products from the primary PCR batch of Fig. 3 as a template;
- Fig. 5 shows a gel retardation experiment;
- 40

Fig. 6 shows another gel retardation experiment;

Fig. 7 shows the elution of calf thymus DNA and pUC18*emm* on rCpG-181 sepharose, and

Fig. 8 shows the determination of the eluted DNA in the fractions by measurement of the extinction at 254 nm as a function of the NaCl gradient.

The wild type CPGB protein CPGbP656 binds non-methylated CpG motifs of prokaryotic DNA, thus forming a protein-DNA complex. This complex may be or become bound to a carrier, for example, whereby separation and/or enrichment of DNA can be effected. Now, the present invention is based on the surprising finding that a protein of the invention which is shortened in comparison with the wild type CPGB protein (CPGbP656 comprising 656 amino acids), in particular CPGbP181 comprising 181 amino acids and presenting a 27.6% homology with the wild type CPGB protein, has improved binding properties with respect to non-methylated CpG motifs of prokaryotic DNA than the wild type CPGB protein and variants thereof with a homology of 80% or more.

Prokaryotic DNA differs from eukaryotic DNA, for example, by the occurrence of non-methylated CpG motifs (Deutsches Ärzteblatt, Jg.[vol.] 98/15: A981-A985 (2001)). The invention is based on the finding that eukaryotic DNA and prokaryotic DNA differ in their proportion of CpG motifs. In prokaryotic DNA, CpG motifs are present with a 20-fold excess as compared to eukaryotic DNA which contains such motifs only temporarily, e.g., in cancer cells or promoter regions (Deutsches Ärzteblatt, Jg. 98/15: A981-A985 (2001)). In prokaryotic DNA, these motifs are not methylated, whereas most of them are methylated in eukaryotic DNA, which additionally increases their distinctiveness. Non-methylated CpG motifs are non-methylated desoxycytidylate-desoxyguanylate dinucleotides in the prokaryotic genome or in fragments thereof.

The invention is further based on the finding that the protein of the invention specifically binds to non-methylated CpG motifs. This specific binding property of the protein of the invention is utilized in order to bind prokaryotic DNA and thus to subsequently enrich it, separate it, and isolate it from a sample including, e.g., a majority of eukaryotic DNA.

The term "homology" within the meaning of the present invention designates the degree of identity of two protein sequences. A homology of x% here means that x out of 100 amino acid positions in the sequences are identical. The term "shortened" as used for characterizing the protein of the invention means that the length of the amino acid sequence of the protein of the invention (CPGbP181) is shorter than the length of the

amino acid sequence of the wild type CPGB protein (CPGbP656). Shortening is effected at the N-terminus and at the C-terminus of the wild type protein sequence (Fig. 1). The maximum shortening is represented by the DNA binding site of the protein. This means that the protein of the invention is shortened, at the most, to the DNA binding site as compared to the wild type protein.

The protein of the invention may preferably have a molecular weight of about 19,959 Dalton (native) or 21,444 Dalton (in the plasmid pQE60). In another preferred embodiment, the isoelectric point of the protein of the invention is about 10.09 (native protein) or 10.15 (in the plasmid pQE60). A particularly preferred protein of the invention has the amino acid sequence shown in SEQ ID No. 2 or in Fig. 1, respectively. This protein has particularly good binding properties as compared to non-methylated CpG motifs of prokaryotic DNA.

The protein described in EP 02020904 (CPGbP241), which is a shortened variant of the wild type CPGB protein (CPGbP656) and served as the basis for the protein employed according to the invention (e.g. CPGbP181), has a length of 241 amino acids, a molecular weight of approximately 33,650 Dalton (native) or 28,138 Dalton (in the plasmid pQE60) and an isoelectric point of 9.89 (native) or 9.88 (in plasmid pQE60). The cDNA and amino acid sequence is shown in Figs. 1 and 2.

The wild type CGBP protein has a length of 656 amino acids, 135 positively charged residues and 94 negatively charged residues, a molecular weight of approximately 75,684 Dalton, and an isoelectric point of 8.15. The cDNA and amino acid sequence is shown in Fig. 1.

The sequence comparison of the protein of the invention (CPGbP181) according to SEQ ID No. 2 with the protein described in EP 01010904 (CPGbP241) is shown in Figs. 1 and 2.

The protein of the invention is preferably produced by cloning the corresponding cDNA sequence into a plasmid and by expression in *Escherichia coli*. An *E.coli* strain expressing the protein of the invention was deposited with the Deutsche Sammlung für Mikroorganismen und Zellkulturen under No. DSM 16229 on Feb. 16, 2004. Alternatively, other production methods known in the art can be applied. The use of the plasmid pQE9 represents an exemplary possibility, but any other suitable plasmid may be employed as a vector. Expression in *E.coli* is also just an example. Expression in other prokaryotic systems and also in a eukaryotic system as well as chemical or enzymatic synthesis or purification from a genetically modified tobacco plant are further possible embodiments of

protein extraction. The protein can be produced both on a laboratory scale (e.g., in an Erlenmeyer flask) and on an industrial scale (e.g., fermenter). For example, the protein of the invention can be purified by binding histidine residues (His-tag), which are introduced to the beginning or to the end of the protein, to a suitable nickel-containing matrix, which is a method known in the art.

Further possibilities of purification may be any type of fusion proteins allowing purification via suitable matrices (columns, gels, beads, etc.). Other forms of tags may be fusion peptides/fusion proteins, e.g., streptavidin tag, Myc tag, and others.

A preferred form of the protein of the invention is the native form, but a denatured form is also suitable for binding non-methylated CpG motifs. Within the meaning of the present invention, "denatured forms" are understood to be secondary structures other than those found in nature.

The native, or denatured, form of the protein of the invention constitutes an exemplary embodiment. The invention includes *in vitro* synthesis as well as any other chemical or enzymatic modifications of the protein such as, e.g., incorporation of disulfide bridges, glycosilations, phosphorylations, acylations, amino acid exchanges, as well as fusion with proteins or other molecules. Such modifications may be achieved, for example, by recombination and/or expression and/or chemical and/or enzymatic modification of single or multiple amino acids.

The protein of the invention exhibits a multiplicity of advantages. It is better capable of binding prokaryotic DNA via non-methylated CpG motifs than the wild type CPGB protein or variants thereof with a homology of 80% or more. This makes it possible to specifically separate and/or enrich the prokaryotic DNA from a mixture of prokaryotic and eukaryotic DNA. This ultimately enables a quick and simple detection of pathogens as well as early diagnosis of infections which may be caused by bacterial pathogens. Conversely, the invention can also be used for depleting microbial DNA in the sense of purification in the case of clinical conditions accompanied by a non-physiological presence of bacteria or their cleavage products in body fluids, in particular blood, of patients. This applies even more because it is well documented that bacteria but also their cleavage products such as, for example, bacterial DNA, are responsible for a multiplicity of biological effects detrimental to the patient.

A further subject matter of the present invention is antibodies directed against proteins of the invention. These may be monoclonal oder polyclonal antibodies. They may be

produced in a manner known per se; the necessary materials and methods are known to the person having skill in the art.

5 The antibodies may be used for the isolation and quantification of the proteins of the invention. These, too, are possible applications that are known per se, with the necessary materials and methods being known to the person having skill in the art.

10 A further subject matter of the invention is nucleic acid, in particular DNA coding for a protein of the invention. Such a DNA is the one represented in SEQ ID No. 1 (or Fig. 2).

Due to the good binding ability of the protein of the invention to non-methylated CpG motifs of prokaryotic DNA, a further subject matter of the invention is a method of separating and/or enriching prokaryotic DNA, comprising the steps of:

- 15 a) contacting at least one prokaryotic DNA present in solution with the protein of the invention, thus forming a protein-DNA complex, and
 b) separation of said complex.

20 This DNA can be purified and dissolved again or may be present directly in the source of origin (e.g. body fluid such as blood, serum, tracheal aspirate, urine, bronchoalveolar lavage, nose smear, skin smear, puncture fluid).

25 Separation may be effected by various methods of separating, isolating or enriching DNA protein complexes or DNA polypeptide complexes that are well-known to the person skilled in the art. In doing so, use will preferably be made of methods in which the DNA-binding protein is or is being immobilized to a carrier in order to separate and/or enrich the DNA from the sample solution.

30 According to a preferred embodiment, separation is followed by a step of separating the DNA from the protein of the invention from the complex. This may be effected, for example, by conventional methods of DNA purification known to the person skilled in the art. In the most simple case, separation is based on changing the pH value or the salt concentration (e.g., to 1 M NaCl) of the medium/buffer or by adding chaotropic reagents, etc.; i.e., suitable parameters which lead to the dissolution of the protein-DNA complex.
35 Such methods are known to the person skilled in the art.

According to a further preferred embodiment, the protein of the invention is bound to a carrier. This embodiment represents a particularly simple way of enriching prokaryotic DNA, for separation from the solution may be effected in a particularly simple manner,

e.g., by physical removal (e.g. by centrifugation) of the charged carrier(s) from the solution.

For the solution for the prokaryotic DNA, any suitable solvent is basically conceivable.

5 However, the method is particularly expedient for enriching prokaryotic DNA from solutions containing different biomolecular species, in particular different kinds of DNA. The invention preferably relates to a method of separating and enriching prokaryotic or viral DNA from a mixture of prokaryotic and eukaryotic DNA. In doing so, for example the prokaryotic DNA present in body fluids is separated from the eukaryotic DNA and
10 enriched by specific binding to the protein of the invention. The prokaryotic DNA thus enriched facilitates the detection of prokaryotic pathogens with the aid of molecular-biological methods and may contribute to the diagnosis of diseases caused by pathogenic germs.

15 Particularly the embodiment in which the DNA-binding protein of the invention is immobilized to the surface of a carrier is suitable for adsorbing prokaryotic DNA from body fluids, preferably from blood. This approach moreover allows to remove microbial DNA, which is present in blood or other body fluids, from these fluids. The body fluid (e.g. whole blood, serum or liquor) purified in this way of the microbial DNA - which by itself is
20 capable of causing severe inflammatory reactions in patients - can then be fed back into the body. This principle may also be used for depleting prokaryotic DNA from physiological fluids in the sense of a purification, wherein the specific binding properties of the protein of the invention are made use of.

25 Body fluids within the meaning of the invention are understood to be any fluids originating from the body of a mammal, including humans, in particular such fluids in which disease pathogens may occur, such as blood, urine, liquor, pleural liquid, pericardial liquid, peritoneal liquid, as well as synovial liquid. The description of the invention referring to human blood is not to be construed as a limitation but only as an exemplary application.

30 Bacterial pathogens are preferably understood to be pathogens of a sepsis, but also any other bacterial pathogens of infections. They may differ from commensal pathogens which are counted among the normal population of the organism and are occasionally also found in test samples from patients but do not have any clinical significance.

35 When isolating total DNA from infected body liquids, the ratio of host DNA to pathogen DNA may in many cases be only $1:10^{-6}$ to $1:10^{-8}$ or even less. Due to the specific binding of prokaryotic DNA to the protein of the invention, the method according to the invention enables enrichment by 1 exponential unit and more.

The protein of the invention may be coupled directly or indirectly to the carrier. The type of coupling depends on the carrier and the carrier material. Suitable carriers include in particular membranes, microparticles and resins or similar materials for affinity matrices.

5 Suitable materials for binding the protein of the invention, as well as - depending on the type of material - for carrying out such binding are well-known to the person skilled in the art. For indirect coupling, for example, specific antibodies against the protein of the invention or the polypeptide are suitable which are in turn bound to the carrier by known methods.

10 One application of the method according to the invention consists in enriching prokaryotic DNA. A further application consists in the separation of prokaryotic DNA from a mixture of eukaryotic and prokaryotic DNA by binding the prokaryotic DNA to the protein of the invention which was immobilized, for example, to a matrix. The mixture of the body's own
15 DNA and prokaryotic DNA is contacted to the affinity matrix by means of suitable methods, and in doing so, the prokaryotic DNA is bound to the immobilized protein of the invention; for example, the eukaryotic DNA passes through a separating column and may be collected separately. Affinity matrices may be, for example, polymeric polysaccharides such as agaroses, other biopolymers, synthetic polymers, or carriers having a silicate
20 backbone such as porous glasses or other solid or flexible carriers on which the DNA-binding protein of the invention is immobilized. After the separation of prokaryotic DNA from eukaryotic DNA has been effected, the affinity matrix is rinsed with a suitable reagent, so that the binding protein with the coupled prokaryotic DNA is separated from the matrix and/or the prokaryotic DNA is separated from the binding protein and is
25 available in a sufficient amount for further process steps.

Another application of the method according to the invention consists in the separation and enrichment of prokaryotic DNA from eukaryotic DNA by binding the prokaryotic DNA to the protein of the invention which was immobilized on microparticles. In this
30 connection, all microparticles which allow an immobilization of the DNA-binding protein of the invention are suitable. Such microparticles may consist of latex, plastics (e.g. styrofoam, polymer), metal, or ferromagnetic substances. Furthermore, use may also be made of fluorescent microparticles such as those available, e.g., from the company Luminex. After the prokaryotic DNA has been bound to the proteins according to the
35 invention which are immobilized on microparticles, said microparticles are separated from the mixture of substances by suitable methods such as filtration, centrifugation, precipitation, sorting by measuring the intensity of fluorescence, or by magnetic methods.

After separation from the microparticles, the prokaryotic DNA is available for further processing.

5 Another application of the method according to the invention consists in the separation and enrichment of prokaryotic DNA from eukaryotic DNA by binding the prokaryotic DNA to the protein of the invention which is subsequently separated from other ingredients of the mixture by electrophoresis.

10 A further application of the method according to the invention consists in the separation and enrichment of prokaryotic DNA from eukaryotic DNA by binding the prokaryotic DNA to the protein of the invention, wherein the protein of the invention is subsequently bound to corresponding antibodies. The antibodies may be bound to solid or flexible substrates such as glass, plastics, silicon, microparticles, membranes, or may be present in solution. After binding of the prokaryotic DNA to the protein of the invention and binding of the
15 latter to the specific antibody, separation from the mixture of substances is effected by methods that are known to the person skilled in the art.

The method according to the invention may also be used for purifying body fluids of prokaryotic DNA. Here it is convenient to perform the separation extracorporeally under
20 sterile conditions, to allow the body fluids to be fed back into the body again, so that the body's own immune system is assisted in eliminating infections by removing the prokaryotic DNA present in said body fluids.

Any suitable chemical, mechanical or electrochemical processes may be considered for
25 the extracorporeal removal of prokaryotic DNA from body fluids. In addition, the combination with other extracorporeal methods such as hemoperfusion, heart-lung machine or endotoxin adsorbers, constitutes another expedient application.

The protein of the invention can also be used to detect prokaryotic DNA. In this case,
30 enrichment of the prokaryotic DNA is followed by a step of amplifying the prokaryotic DNA, for which any common amplification methods are suitable (PCR, LCR; LM-PCR, etc.).

The invention moreover relates to a kit for enriching prokaryotic DNA by means of one of
35 the above-described methods, said kit containing at least the protein of the invention, optionally together with further reagents suitable for carrying out said method.

In addition to the protein of the invention, said kit may contain at least one set of primers which are suitable for amplifying genomic DNA of certain prokaryotes under standard conditions.

5 The method according to the invention, in particular with the above-described embodiments, has the advantage that, by specific binding of non-methylated prokaryotic DNA rich in CpG motifs to proteins with specific affinity for such structures, prokaryotic DNA from the total DNA of an infected host is successfully concentrated and thus the sensitivity of detection of pathogen DNA in body fluids is enhanced highly.

10 The possibilities of separating prokaryotic DNA from eukaryotic DNA by means of a specifically binding protein are no more time-consuming than known methods of isolating total DNA. Subsequent DNA detection may be performed by way of a PCR reaction. In most cases a nested PCR will not be necessary, which makes it possible to save a considerable amount of time in diagnostics.

The use of the protein of the invention in order to deplete prokaryotic DNA in physiological body fluids was already mentioned above. Depletion within the meaning of the present invention means that the quantity of prokaryotic DNA is reduced. This possibility of reducing prokaryotic DNA also enables the use of the proteins according to the invention in environmental technology, waste water management, and air conditioning technology.

25 In the following, the invention will be explained in more detail by way of the examples, however without limiting it thereto.

Example 1: Preparation of the protein of the invention

30 The DNA sequence for the complete CPGBP protein was used to construct primer 1 (GGATCCGGTGGAGGGCGCAAGAGGCCTG —fw SEQ ID No. 3) and primer 2 (AAGCTTAGAGGTAGGTCCTCAT-CTGAG-rv SEQ ID No. 4) which amplify a shortened DNA fragment that codes for a shortened CPGB-binding protein, CPGBP-181. After cleavage, the DNA fragment was ligated into the vector pQE9 (Qiagen) by using the restriction enzymes *Bam*HI and *Hind* III. An open reading frame forms in pQE9, in which frame a DNA fragment coding 6* His-Tag (pQE9[6HisCPGBP181]) is fused to the 5' end.

Plasmid pQE9[6HisCPGBP181] was transformed to the *E. coli* expression strain M15[pREP4] (Qiagen). The clone shall in the following be referred to as

M15[pCPGbP181], and the expressed protein as rCPGbP181. Expression of the protein rCPGbP181 took place according to the following protocol: A colony of the expression strain M15[pCPGbP181] is grown/angezüchtet overnight in 2 ml of Luria Medium with 100 µg/ml of ampicilline and 25 µg/ml of kanamycine at 37°C with shaking. Then, the pre-
5 culture is transferred to 200 ml of preheated nutrient medium containing the same concentrations of antibiotics. After 3 hours of growth at 37°C with shaking, IPTG is added to induce expression, and incubation is continued for 5 hours. Thereafter, the bacteria are removed by centrifugation and the sediment is re-suspended in 5 ml of 0.2 M tris buffer, pH 7.5. The bacteria are subjected to ultrasonic treatment in an iced bath for 5*1 min.

10 After centrifugation, the sediment is re-suspended in 10 ml of 0.2 M tris, 2M urea, pH 7.5, and shaken for 15 min. After centrifugation has been effected, the remaining sediment is taken up and suspended in 0.2 M tris, 6M guanidine hydrochloride, 0.001 M dithioerythritol (DTE), 0.02 M imidazole. The inclusion bodies are dissolved at room temperature for 1 hour with agitation. After centrifugation, the crude protein is present in the supernatant
15 and can be applied directly to a 3 ml Ni-agarose column. The subsequent steps should take place in the cooling chamber at +4 to +6°C. First, the column is washed with 0.2 M tris, 6M guanidine hydrochloride, 0.001 M dithioerythritol (DTE), 0.02 M imidazole buffer, pH 7.5, until extinction has reached the zero line. From this point, rCPGbP181 can be obtained in different ways: 1. as a denatured protein, dissolved in 6M guanidine
20 hydrochloride or 6M urea, and 2. as a native protein, soluble in buffers at physiological concentrations. In the second case, however, the yield is lower.

Purification according to method 1 (denatured):

The protein rCPGbP181 is eluted from Ni-NTA agarose with an imidazole gradient of 0-
25 0.5 M, M in buffer 0.2 M tris, 6 M guanidine hydrochloride, 0.001 M dithioerythritol (DTE), 0.02 M imidazole, pH 7.5, as the basic material. In doing so, rCPGbP181 is detached from the column at 0.2-0.3 M imidazole. The protein thus obtained is dialyzed against 0.2 M tris, 6M urea, 0.001 M dithioerythritol (DTE), pH 7.5, and frozen. During dialysis against physiological buffers, purified rCPGbP181 is thus precipitated.

30 Purification according to method 2 (native):

According to this method, the guanidine hydrochloride concentration is shifted from 6 mol on Ni-NTA agarose with the bound rCPGbP181 via a gradient to 0 mol guanidine hydrochloride. The basis for this is the buffer 0.2 M tris, 0.5 M NaCl, 0.001 M
35 dithioerythritol (DTE), 0.02 M imidazole, pH 7.5. The flow rate was 0.5 ml/min.

Subsequently, an imidazole gradient of 0 to 0.5 mol was applied for elution in buffer 0.2 M tris, 0.5 M NaCl, 0.001 M dithioerythritol (DTE), pH 7.5, as basic material. In this case, too, a substantial proportion of the bound protein (20%) was eluted at 0.2 to 0.3 mol

imidazole. This native rCPGbP181 eluate remained dissolved in this buffer even after dialysis in PBS. However, it is disadvantageous that under these conditions approximately 80% of the rCPGbP181 bound to Ni-NTA agarose remained on the column and could subsequently only be extracted under the denaturing conditions of method 1.

- 5 This means that the yield of method 2 as used resulted only in 20% native rCPGbP181 soluble in physiological buffers.

Example 2: Detection of pathogens by means of nested PCR:

- 10 Fresh heparinized human blood containing *Streptococcus pyogenes* with 10^3 /ml colony-forming units as the pathogen is used for pathogen detection. The DNA is isolated by means of absorption to DNA-binding matrix by using commercial kits for the isolation of total DNA from body fluids according to modified instructions from the manufacturers. For this purpose, 200 µl of the total lysis buffer, which contains proteinase K and SDS, is
- 15 added to 100 µl of infected blood in Eppendorf tubes. The mixture is incubated at 37°C for 30 min and then heated to 95°C for 20 min. After cooling, 20 µg of mutanolysine is added and incubation at 37°C is continued for another 60 min. After centrifugation, the mixture is applied to the centrifugation columns using DNA-binding matrix, and the DNA is purified according to manufacturer's instructions. The purified DNA is placed in a final
- 20 volume of 100 µl of 0.01 mol tris buffer, pH 7.5, or in an equal amount of elution buffer from the manufacturer. For detection of pathogens, primers were selected to identify the streptolysin O gene (slo).

1. PCR. Amplification of a 465 bp fragment

- 25 Forward primer 1: 5'-AGCATACAAGCAAATTTTTTACACCG (SEQ ID No. 5)

Reverse primer 2: 5'-GTTCTGTTATTGACACCCGCAATT (SEQ ID No. 6)

Primer concentration 1 mg/ml

Starting material: 5 µl isolated DNA

- 30 0.5 µl primer fw 1

0.5 µl primer rv 2

14 µl aqua dest

total 25 µl in Ready to go Kit (Amersham-Pharmacia)

- 35 Reaction:

5 min 95°C

40 cycles (30 sec. 95°C; 30 sec. 51°C; 3 min 72°C; 1*7 min 72°C).

The first PCR of streptococci DNA in human blood is shown in Fig. 1 (10 µl each of the 25 µl starting material were separated. 1) PCR starting material containing 5 µl of template DNA; 2) starting material containing 5 µl of template, at a dilution of 1:10. 3) positive control: 0.2 µl of streptococci-DNA as template in the absence of eucaryotic DNA from blood. ST) molecular weight standard)

Result: The first primary PCR does not result in a positive reaction. Therefore, a second PCR (nested PCR) was subsequently carried out.

2. Nested PCR: Amplification of a 348 bp fragment in the above *s/o*-fragment.

Forward primer 3: 5'-CCTTCCTAATAATCCTGCGGATGT (SEQ ID No. 7)

Reverse primer 4: 5'-CTGAAGGTAGCATTAG TCTTTGATAACG (SEQ ID No. 8)

Primer concentration: 1 mg/ml

Starting material: 5 µl from PCR1, sample 1, Fig. 3
0.5 µl primer fw 1
0.5 µl primer rv 2
14 µl aqua dest
total 25 µl in Ready to go Kit (Amersham-Pharmacia)

Reaction:

5 min 95°C

40 cycles (30 sec. 95°C; 30 sec. 54°C; 3 min 72°C; 1*7 min 72°C)

Fig. 4 shows the nested PCR with the PCR products from the primary PCR starting material according to Fig. 3 as the template. The samples correspond to those of Fig. 3.

Result: In the nested PCR, the desired *s/o*-DNA fragment is amplified at a concentration of 100 streptococci cells per 100 µl of blood (sample 1). For 5 µl starting material in the 1st PCR (Fig. 3), this corresponds to about 5 to 10 templates. At a dilution of 1:10 (sample 2), sensitivity is exhausted (0.5 to 1 template).

These experiments show that successful PCR detection of pathogens in blood requires isolation of the total DNA from at least 1 to 5 ml of blood. However, the total DNA concentration is then too massive for being used directly in a PCR.

Other pathogen-specific nucleic acid detections without an amplification step by direct detection of the bacterial DNA, for example by DNA hybridization, are also too insensitive, which is primarily due to the high excess of human DNA relative to bacterial DNA. In

addition, competitive processes during DNA analysis as well as the low quantity of bacterial DNA are to be regarded as hindrances to qualitative and quantitative analysis. The common methods of DNA isolation enrich the total DNA of a body fluid so that the ratio of host DNA to microbial DNA can be between $1:10^{-6}$ and $1:10^{-8}$. This difference makes it easy to understand the difficulty in detecting microbial DNA in body fluids.

Example 3: Determining the binding properties of rCPGbP181:

In gel retardation experiments both the binding of the denatured and of the native protein rCpGbP181 to methylated and to non-methylated DNA molecules with CpG motifs was examined. The pUC18 plasmid of *E. coli* was used as the test DNA with an inserted M-protein gene segment of *streptococcus dysgalactiae subsp. equisimilis* (Geyer et. al FEMS Immuno. Med. Microbiol. 26:11-24, 1999). The plasmid preparation was divided and one half was methylated with the CpG methylase kit of New England BioLabs. Both preparations were mixed with rCPGbP181 (native or denatured) and electrophoretically separated on agarose gel. The results are shown in Figs. 5 and 6. Both the native form and the denatured form of rCPGbP181 showed a higher affinity to non-methylated plasmid DNA, which confirms the selective binding property with respect to non-methylated CpG-rich DNA.

Description of the gel retardation experiment according to Fig. 5: 5 µl (72 ng) methylated pUC18*emm* DNA and 1 µl (142 ng) non-methylated pUC18*emm* DNA, respectively, were mixed with 5 µl (0.5 µg) native rCPGbP181 and filled up to a volume of 35 µl with the following buffer: 0.01 M tris, 0.08M NaCl, 0.001M EDTA, 0.005M DTE, 5% glycerine, pH 7.8. After incubation at 20°C for 30 min the mixtures were electrophoretically separated on 1.5% agarose. Methylated DNA was applied in lanes 1 and 3 and non-methylated DNA was applied in lanes 2 and 4. In lanes 1 and 2 the DNA was mixed with native rCPGbP181. Lane 2 shows that non-methylated pUC18*emm* interacts with rCPGbP181; in contrast thereto, rCPGbP181 did not show any interaction with methylated pUC18*emm* (lane 1). Lanes 4 and 5 are the plasmids without addition of rCPGbP181 as controls.

Description of the gel retardation experiment of Fig. 6 for non-methylated and methylated pUC18*emm* after incubation with denatured rCPGbP181. The concentrations correspond to those of Fig. 5. Methylated DNA was applied in lanes 1 and 3 and non-methylated DNA was applied in lanes 2 and 4. In lanes 1 to 4, the DNA was mixed with two different batches of denatured rCPGbP181. Lanes 2 and 4 show that non-methylated pUC18*emm* also interact with denatured rCPGbP181; however, rCPGbP181 did not show any

interaction with methylated pUC18*emm* (lanes 1 and 3). Lane 5 is pUC18*emm* without rCPGbp181 as a control.

Example 4: Binding and separation of a mixture of calf thymus DNA and bacterial DNA on immobilized CPGbp181

Purified CPGbp181 was coupled to aminohexyl sepharose (Amersham-Biosciences) by means of glutaraldehyde according to the protocol of Cambiasso et al. (Cambiasso, C. et al., *Immunochemistry* 12-273-278, 1975). The concentration of immobilized protein was 0.3 mg per milliliter sepharose. 300 µl of sepharose was placed in a spin-filter tube containing inert fritting material which absorbs neither DNA nor protein, but retains sepharose.

200 ng of calf thymus DNA and 25 ng pUC18*emm* was dissolved in 100 µl to 20 mM tris-HCL buffer, pH 7.5, and applied to the column thus prepared. After each step, the liquid was centrifuged at 14,000 RPM for 0.5 min in an Eppendorf centrifuge in one fresh Eppendorf tube each. Thus, the NaCl concentration was increased in 50 mM-steps from 0 to 1mM. DNA precipitation was effected in each tube by adding 10 µl of 4 M acetate, pH 4.5, and 250 µl of ethanol abs., mixing and centrifugation at 14,000 RPM for 15 min. Thereafter, the supernatant was discarded and the precipitate was washed with 300 µl of 70% ethanol. After discarding, the residue was dried for 5 min in a vacuum centrifuge and then taken up in 15 l distilled water (PCR-suitable). On the one hand, extinction at 254 nm was measured for 10 µl each of the samples (Fig. 7). On the other hand, PCR was effected with sequence primers for PUC18, using 3 µl of each sample (Fig. 8).

The result (Figs. 7, 8) shows that the eukaryotic calf thymus DNA is initially washed from the column between 0 to 0.1 M NaCl, while the prokaryotic DNA (pUC18*emm*) was eluted in the fraction at 0.3 M NaCl. This shows that eukaryotic DNA has a lower affinity to CPGbp181 and thus, a clear separation of both DNA fractions was achieved.

SEQUENCE LISTING

<110> SIRS-Lab GmbH

5 <120> Protein having improved binding properties for prokaryotic DNA and method for separating and enriching prokaryotic DNA

10 <130> Pat 3696/29

<140>

<141>

15 <160> 8

<170> PatentIn Ver. 2.1

<210> 1

20 <211> 543

<212> DNA

<213> Homo sapiens

<220>

25 <221> CDS

<222> (1)..(561)

<400> 1

30 ggt gga ggg cgc aag agg cct gtc cct gat cca aac ctg cag cgc cgg 48
Gly Gly Gly Arg Lys Arg Pro Val Pro Asp Pro Asn Leu Gln Arg Arg
1 5 10 15

gca ggg tca ggg aca ggg gtt ggg gcc atg ctt gct cgg ggc tct gct 96
Ala Gly Ser Gly Thr Gly Val Gly Ala Met Leu Ala Arg Gly Ser Ala
35 20 25 30

tcg ccc cac aaa tcc tct ccg cag ccc ttg gtg gcc aca ccc agc cag 144
Ser Pro His Lys Ser Ser Pro Gln Pro Leu Val Ala Thr Pro Ser Gln
35 40 45

40 cat cac cag cag cag cag cag cag atc aaa cgg tca gcc cgc atg tgt 192
His His Gln Gln Gln Gln Gln Gln Ile Lys Arg Ser Ala Arg Met Cys
50 55 60

45 ggt gag tgt gag gca tgt cgg cgc act gag gac tgt ggt cac tgt gat 240
Gly Glu Cys Glu Ala Cys Arg Arg Thr Glu Asp Cys Gly His Cys Asp
65 70 75 80

50 ttc tgt cgg gac atg aag aag ttc ggg ggc ccc aac aag atc cgg cag 288
Phe Cys Arg Asp Met Lys Lys Phe Gly Gly Pro Asn Lys Ile Arg Gln
85 90 95

aag tgc cgg ctg cgc cag tgc cag ctg cgg gcc cgg gaa tcg tac aag 336
Lys Cys Arg Leu Arg Gln Cys Gln Leu Arg Ala Arg Glu Ser Tyr Lys
55 100 105 110

tac ttc cct tcc tcg ctc tca cca gtg acg ccc tca gag tcc ctg cca 384
Tyr Phe Pro Ser Ser Leu Ser Pro Val Thr Pro Ser Glu Ser Leu Pro

	115	120	125	
5	agg ccc cgc cgg cca ctg ccc acc caa cag cag cca cag cca tca cag	432		
	Arg Pro Arg Arg Pro Leu Pro Thr Gln Gln Gln Pro Gln Pro Ser Gln			
	130	135	140	
10	aag tta ggg cgc atc cgt gaa gat gag ggg gca gtg gcg tca tca aca	480		
	Lys Leu Gly Arg Ile Arg Glu Asp Glu Gly Ala Val Ala Ser Ser Thr			
	145	150	155	160
15	gtc aag gag cct cct gag gct aca gcc aca cct gag cca ctc tca gat	528		
	Val Lys Glu Pro Pro Glu Ala Thr Ala Thr Pro Glu Pro Leu Ser Asp			
	165	170	175	
20	gag gac cta cct ctg	543		
	Glu Asp Leu Pro Leu			
	180			
25	<210> 2			
	<211> 181			
	<212> PRT			
	<213> Homo sapiens			
30	<400> 2			
	Gly Gly Gly Arg Lys Arg Pro Val Pro Asp Pro Asn Leu Gln Arg Arg			
	1	5	10	15
35	Ala Gly Ser Gly Thr Gly Val Gly Ala Met Leu Ala Arg Gly Ser Ala			
	20	25	30	
40	Ser Pro His Lys Ser Ser Pro Gln Pro Leu Val Ala Thr Pro Ser Gln			
	35	40	45	
45	His His Gln Gln Gln Gln Gln Gln Ile Lys Arg Ser Ala Arg Met Cys			
	50	55	60	
50	Gly Glu Cys Glu Ala Cys Arg Arg Thr Glu Asp Cys Gly His Cys Asp			
	65	70	75	80
55	Phe Cys Arg Asp Met Lys Lys Phe Gly Gly Pro Asn Lys Ile Arg Gln			
	85	90	95	
60	Lys Cys Arg Leu Arg Gln Cys Gln Leu Arg Ala Arg Glu Ser Tyr Lys			
	100	105	110	
65	Tyr Phe Pro Ser Ser Leu Ser Pro Val Thr Pro Ser Glu Ser Leu Pro			
	115	120	125	
70	Arg Pro Arg Arg Pro Leu Pro Thr Gln Gln Gln Pro Gln Pro Ser Gln			
	130	135	140	
75	Lys Leu Gly Arg Ile Arg Glu Asp Glu Gly Ala Val Ala Ser Ser Thr			
	145	150	155	160
80	Val Lys Glu Pro Pro Glu Ala Thr Ala Thr Pro Glu Pro Leu Ser Asp			
	165	170	175	

Glu Asp Leu Pro Leu
180

5
 <210> 3
 <211> 28
 <212> DNA
 <213> Artificial sequence
10
 <220>
 <223> Description of the artificial sequence: Primer

 <400> 3
15 ggatccggtg gagggcgcaa gaggcctg 28

 <210> 4
 <211> 27
20 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Description of the artificial sequence: Primer
25
 <400> 4
 aagcttagag gtaggtcctc atctgag 27

30 <210> 5
 <211> 26
 <212> DNA
 <213> Artificial sequence

35 <220>
 <223> Description of the artificial sequence: Primer

 <400> 5
40 agcatacaag caaatttttt acaccg 26

 <210> 6
 <211> 24
 <212> DNA
45 <213> Artificial sequence

 <220>
 <223> Description of the artificial sequence: Primer

50 <400> 6
 gttctgttat tgacacccgc aatt 24

55 <210> 7
 <211> 24
 <212> DNA
 <213> Artificial sequence

<220>

<223> Description of the artificial sequence: Primer

<400> 7

5 ccttcctaat aatcctgcgg atgt

<210> 8

<211> 28

10 <212> DNA

<213> Artificial sequence

<220>

<223> Description of the artificial sequence: Primer

15

<400> 8

ctgaaggtag cattagtctt tgataacg

20

SIRS-Lab GmbH
Attorney's file: PAT 3696/029

March 5, 2004
H/18/kt

Claims

5

1. A protein which binds non-methylated CpG motifs while having a 25% to 35% homology with the wild type CGPB protein and being shorted in comparison with the latter, wherein the binding site for non-methylated CpG motifs is contained.

10

2. The protein according to any one of the preceding claims, wherein it includes the amino acid sequence in accordance with SEQ ID No. 2.

15

3. The protein according to any one of the preceding claims, wherein it is produced by modification.

20

4. The protein according to claim 3, wherein the modification is achieved by recombination and/or expression and/or chemical and/or enzymatic modification of single or several amino acids.

25

5. The protein according to claim 4, wherein the modification is achieved by the incorporation of disulfide bridges, glycosilations, phosphorylations, acylations, amino acid exchanges, as well as fusion with further proteins or other molecules.

6. The protein according to any one of the preceding claims, wherein it is shortened to the length of the binding site at the most.

30

7. An antibody against the proteins defined in accordance with the preceding claims, wherein the antibodies are monoclonal oder polyclonal antibodies.

8. Use of the antibodies according to claim 7 for isolating and quantifying the protein as defined in accordance with claims 1 to 6.

9. A method of separating and/or enriching prokaryotic DNA, comprising the steps of:

- a. contacting at least one prokaryotic DNA, present in solution, with a protein which specifically binds prokaryotic DNA and has 25% to 35% homology with the wild type CPGB protein, thereby forming a protein-DNA complex, and
- b. separation of said complex.

5

10. The method according to claim 9, wherein separation is followed by a step for separating the DNA from the protein of the complex.

11. The method according to claim 9 oder 10, wherein the protein is bound to a carrier.

10

12. The method according to claim 11, wherein the protein is bound directly to the carrier.

13. The method according to claim 11, wherein the protein is bound to the carrier via an antibody directed against it.

15

14. The method according to any one of claims 11 to 13, wherein the carrier is provided as a matrix, as microparticles, or as a membrane.

20

15. The method according to any one of claims 9 to 14, wherein separation is effected by means of an antibody or antiserum directed against the protein.

16. The method according to any one of claims 9 to 14, wherein separation is effected by means of electrophoresis.

25

17. The method according to any one of claims 12 to 16, wherein the protein is an antibody directed against non-methylated CpG motifs or a corresponding antiserum.

18. The method according to any one of claims 9 to 17, wherein the solution contains a mixture of eukaryotic and prokaryotic DNA.

30

19. The method according to claim 18, wherein the solution is a body fluid or is derived therefrom, in particular full blood, serum, plasma, cell preparations from full blood, urine, liquor, pleural liquid, pericardial liquid, peritoneal liquid, synovial liquid, or bronchoalveolar lavage.

35

20. The method according to any one of claims 17 to 19, wherein separation is achieved by means of a filter which filters out corresponding DNA-protein complexes.

21. The method according to claim 20, wherein the protein is immobilized to a filter matrix.
- 5 22. The method according to any one of claims 9 to 21 for use in environmental technology, water management and waste water management as well as in air conditioning technology.
- 10 23. The method according to any one of claims 9 to 21, wherein as a step c) after step b), the prokaryotic DNA is furthermore amplified.
24. The method according to claim 23, comprising the steps of:
- 15 a) isolating the prokaryotic DNA from the protein-DNA complex,
b) denaturing the double-stranded DNA,
c) hybridizing the individual strands of the DNA with complementary primers,
d) generating double-strand fragments via reaction with polymerases, and
e) repeating these steps up to the desired degree of amplification.
- 20 25. The method according to claim 24, comprising the steps of:
- 25 a) cloning the isolated prokaryotic DNA sequences into vectors,
b) transforming suitable host cells with these vectors,
c) cultivating these transformed cells,
d) isolating the vectors from these cells, and
e) isolating the DNA.
- 30 26. A kit for enriching prokaryotic DNA by means of a method according to any one of claims 9 to 25.
27. A test kit for detection of prokaryotic DNA by means of a method according to any one of claims 9 to 25, using one or several sets of specific primers.
- 35 28. Use of the proteins according to claims 1 to 6 for the screening of substance libraries with regard to their properties of binding to protein-bound DNA sequences.
29. Nucleic acid which codes for a protein according to any one of claims 1 to 6.

SIRS-Lab GmbH
Attorney's file: PAT 3696/029

March 5, 2004
H/18/kt

Abstract

5

What is disclosed is a protein which binds non-methylated CpG motifs while having a 25% to 35% homology with the wild type CGPB protein and being shorter in comparison with the latter, wherein the binding site for non-methylated CpG motifs is contained. The application moreover relates to a method of separating and/or enriching prokaryotic DNA,
10 comprising the steps of a) contacting at least one prokaryotic DNA, present in solution, with a protein of the invention, thereby forming a protein-DNA complex, and b) separation of said complex.

1/10

		10	20	30	40	50
CPGbP656	1	MEGDGSDPEP	PDAGEDSKSE	NGENAPIYCI	CRKPDINCFM	IGCDNCNEWF
CPGbP241	1	-----	-----	-----	-----	-----
CPGbP181	1	-----	-----	-----	-----	-----
		60	70	80	90	100
CPGbP656	51	HGDCIRITEK	MAKAIREWYC	RECREKDPKL	EIRYRHKKS	ERDGNERNSS
CPGbP241	51	-----	-----	-----	-----	-----
CPGbP181	51	-----	-----	-----	-----	-----
		110	120	130	140	150
CPGbP656	101	EPRDEGGGRK	RPVPDPNLQR	RAGSGTGVGA	MLARGSASPH	KSSPQPLVAT
CPGbP241	101	-----GGGRK	RPVPDPNLQR	RAGSGTGVGA	MLARGSASPH	KSSPQPLVAT
CPGbP181	101	-----GGGRK	RPVPDPNLQR	RAGSGTGVGA	MLARGSASPH	KSSPQPLVAT
		160	170	180	190	200
CPGbP656	151	PSQHHQQQQQ	QIKRSARMCG	ECEACRRTED	CGHCDFCRDM	KKFGGPNKIR
CPGbP241	151	PSQHHQQQQQ	QIKRSARMCG	ECEACRRTED	CGHCDFCRDM	KKFGGPNKIR
CPGbP181	151	PSQHHQQQQQ	QIKRSARMCG	ECEACRRTED	CGHCDFCRDM	KKFGGPNKIR
		210	220	230	240	250
CPGbP656	201	QKCRRLQCQL	RARESYKYFP	SSLSPVTPSE	SLPRPRRPLP	TQQQPQPSQK
CPGbP241	201	QKCRRLQCQL	RARESYKYFP	SSLSPVTPSE	SLPRPRRPLP	TQQQPQPSQK
CPGbP181	201	QKCRRLQCQL	RARESYKYFP	SSLSPVTPSE	SLPRPRRPLP	TQQQPQPSQK
		260	270	280	290	300
CPGbP656	251	LGRIREDEGA	VASSTVKEPP	EATATPEPLS	DEDLPLDPL	YQDFCAGAFD
CPGbP241	251	LGRIREDEGA	VASSTVKEPP	EATATPEPLS	DEDLPLDPL	YQDFCAGAFD
CPGbP181	251	LGRIREDEGA	VASSTVKEPP	EATATPEPLS	DEDLPL----	-----
		310	320	330	340	350
CPGbP656	301	DNGLPWMSDT	EESPFLDPAL	RKRAVKVKHV	KRREKKSEKK	KEERYKRRHQ
CPGbP241	301	DNGLPWMSDT	EESPFLDPAL	RKRAVKVKHV	KRREKKSEKK	KEERYK----
		360	370	380	390	400
CPGbP656	351	KQKHDKWKH	PERADAKDPA	SLPQCLGPGC	VRPAQPSSKY	CSDDCGMKLA
		410	420	430	440	450
CPGbP656	401	ANRIYEILPQ	RIQQWQQSPC	IAEEHGKKLL	ERIRREQQSA	RTRLQEMERR
		460	470	480	490	500
CPGbP656	451	FHELEAILR	AKQOAVREDE	ESNEGSDDDT	DLQIFCVSCG	HPINPRVALR
		510	520	530	540	550
CPGbP656	501	HMERCYAKYE	SQTSFGSMYP	TRIEGATRLF	CDVYNPQSKT	YCKRLQVLCP
		560	570	580	590	600
CPGbP656	551	EHSRDPKVPA	DEVCGCPLVR	DVFELTGDFC	RLPKRQCNRH	YCWEKLRAE
		610	620	630	640	650
CPGbP656	601	VDLERVRVWY	KLDELFEQER	NVRTAMTNRA	GLLALMLHQT	IQHDPLTTDL
		660	670	680	690	700
CPGbP656	651	RSSADR....

Fig. 1

2/10

Fig. 2

5'	ATG	GAG	GGA	GAT	GGT	TCA	GAC	CCA	GAG	CCT	CCA	GAT	GCC	GGG	GAG	GAC	AGC	AAG
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	M	E	G	D	G	S	D	P	E	P	P	D	A	G	E	D	S	K
	TCC	GAG	AAT	GGG	GAG	AAT	GCG	CCC	ATC	TAC	TGC	ATC	TGC	CGC	AAA	CCG	GAC	ATC
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	S	E	N	G	E	N	A	P	I	Y	C	I	C	R	K	P	D	I
	AAC	TGC	TTC	ATG	ATC	GGG	TGT	GAC	AAC	TGC	AAT	GAG	TGG	TTC	CAT	GGG	GAC	TGC
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	N	C	F	M	I	G	C	D	N	C	N	E	W	F	H	G	D	C
	ATC	CGG	ATC	ACT	GAG	AAG	ATG	GCC	AAG	GCC	ATC	CGG	GAG	TGG	TAC	TGT	CGG	GAG
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	I	R	I	T	E	K	M	A	K	A	I	R	E	W	Y	C	R	E
	TGC	AGA	GAG	AAA	GAC	CCC	AAG	CTA	GAG	ATT	CGC	TAT	CGG	CAC	AAG	AAG	TCA	CGG
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	C	R	E	K	D	P	K	L	E	I	R	Y	R	H	K	K	S	R
	GAG	CGG	GAT	GGC	AAT	GAG	CGG	GAC	AGC	AGT	GAG	CCC	CGG	GAT	GAG	GGT	GGA	GGG
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	E	R	D	G	N	E	R	D	S	S	E	P	R	D	E	G	G	G
																G	G	G
	CGC	AAG	AGG	CCT	GTC	CCT	GAT	CCA	AAC	CTG	CAG	CGC	CGG	GCA	GGG	TCA	GGG	ACA
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	R	K	R	P	V	P	D	P	N	L	Q	R	R	A	G	S	G	T
	R	K	R	P	V	P	D	P	N	L	Q	R	R	A	G	S	G	T
	GGG	GTT	GGG	GCC	ATG	CTT	GCT	CGG	GGC	TCT	GCT	TCG	CCC	CAC	AAA	TCC	TCT	CCG
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	G	V	G	A	M	L	A	R	G	S	A	S	P	H	K	S	S	P
	G	V	G	A	M	L	A	R	G	S	A	S	P	H	K	S	S	P
	CAG	CCC	TTG	GTG	GCC	ACA	CCC	AGC	CAG	CAT	CAC	CAG	CAG	CAG	CAG	CAG	CAG	ATC
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	Q	P	L	V	A	T	P	S	Q	H	H	Q	Q	Q	Q	Q	Q	I
	Q	P	L	V	A	T	P	S	Q	H	H	Q	Q	Q	Q	Q	Q	I
	AAA	CGG	TCA	GCC	CGC	ATG	TGT	GGT	GAG	TGT	GAG	GCA	TGT	CGG	CGC	ACT	GAG	GAC
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	K	R	S	A	R	M	C	G	E	C	E	A	C	R	R	T	E	D
	K	R	S	A	R	M	C	G	E	C	E	A	C	R	R	T	E	D

Fig. 2 (continued)

		549			558			567			576			585			594
TGT	GGT	CAC	TGT	GAT	TTC	TGT	CGG	GAC	ATG	AAG	AAG	TTC	GGG	GGC	CCC	AAC	AAG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
C	G	H	C	D	F	C	R	D	M	K	K	F	G	G	P	N	K
C	G	H	C	D	F	C	R	D	M	K	K	F	G	G	P	N	K
		603			612			621			630			639			648
ATC	CGG	CAG	AAG	TGC	CGG	CTG	CGC	CAG	TGC	CAG	CTG	CGG	GCC	CGG	GAA	TCG	TAC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
I	R	Q	K	C	R	L	R	Q	C	Q	L	R	A	R	E	S	Y
I	R	Q	K	C	R	L	R	Q	C	Q	L	R	A	R	E	S	Y
		657			666			675			684			693			702
AAG	TAC	TTC	CCT	TCC	TCG	CTC	TCA	CCA	GTG	ACG	CCC	TCA	GAG	TCC	CTG	CCA	AGG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
K	Y	F	P	S	S	L	S	P	V	T	P	S	E	S	L	P	R
K	Y	F	P	S	S	L	S	P	V	T	P	S	E	S	L	P	R
		711			720			729			738			747			756
CCC	CGC	CGG	CCA	CTG	CCC	ACC	CAA	CAG	CAG	CCA	CAG	CCA	TCA	CAG	AAG	TTA	GGG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
P	R	R	P	L	P	T	Q	Q	Q	P	Q	P	S	Q	K	L	G
P	R	R	P	L	P	T	Q	Q	Q	P	Q	P	S	Q	K	L	G
		765			774			783			792			801			810
CGC	ATC	CGT	GAA	GAT	GAG	GGG	GCA	GTG	GCG	TCA	TCA	ACA	GTC	AAG	GAG	CCT	CCT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
R	I	R	E	D	E	G	A	V	A	S	S	T	V	K	E	P	P
R	I	R	E	D	E	G	A	V	A	S	S	T	V	K	E	P	P
		819			828			837			846			855			864
GAG	GCT	ACA	GCC	ACA	CCT	GAG	CCA	CTC	TCA	GAT	GAG	GAC	CTA	CCT	CTG	GAT	CCT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
E	A	T	A	T	P	E	P	L	S	D	E	D	L	P	L	D	P
E	A	T	A	T	P	E	P	L	S	D	E	D	L	P	L		
		873			882			891			900			909			918
GAC	CTG	TAT	CAG	GAC	TTC	TGT	GCA	GGG	GCC	TTT	GAT	GAC	AAT	GGC	CTG	CCC	TGG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
D	L	Y	Q	D	F	C	A	G	A	F	D	D	N	G	L	P	W
		927			936			945			954			963			972
ATG	AGC	GAC	ACA	GAA	GAG	TCC	CCA	TTC	CTG	GAC	CCC	GCG	CTG	CGG	AAG	AGG	GCA
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
M	S	D	T	E	E	S	P	F	L	D	P	A	L	R	K	R	A
		981			990			999			1008			1017			1026
GTG	AAA	GTG	AAG	CAT	GTG	AAG	CGT	CGG	GAG	AAG	AAG	TCT	GAG	AAG	AAG	AAG	GAG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
V	K	V	K	H	V	K	R	R	E	K	K	S	E	K	K	K	E
		1035			1044			1053			1062			1071			1080
GAG	CGA	TAC	AAG	CGG	CAT	CGG	CAG	AAG	CAG	AAG	CAC	AAG	GAT	AAA	TGG	AAA	CAC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
E	R	Y	K	R	H	R	Q	K	Q	K	H	K	D	K	W	K	H
		1089			1098			1107			1116			1125			1134
CCA	GAG	AGG	GCT	GAT	GCC	AAG	GAC	CCT	GCG	TCA	CTG	CCC	CAG	TGC	CTG	GGG	CCC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
P	E	R	A	D	A	K	D	P	A	S	L	P	Q	C	L	G	P

Fig. 2 (continued)

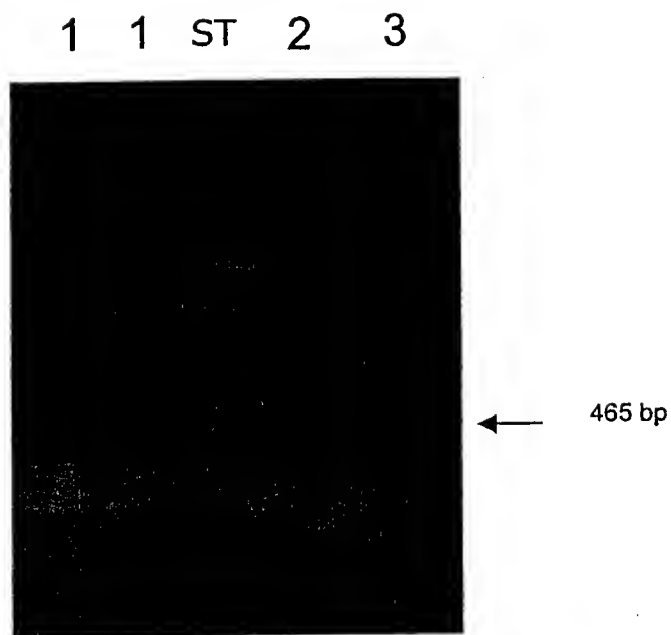
	1143			1152			1161			1170			1179			1188		
GGC	TGT	GTG	CGC	CCC	GCC	CAG	CCC	AGC	TCC	AAG	TAT	TGC	TCA	GAT	GAC	TGT	GGC	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
G	C	V	R	P	A	Q	P	S	S	K	Y	C	S	D	D	C	G	
	1197			1206			1215			1224			1233			1242		
ATG	AAG	CTG	GCA	GCC	AAC	CGC	ATC	TAC	GAG	ATC	CTC	CCC	CAG	CGC	ATC	CAG	CAG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
M	K	L	A	A	N	R	I	Y	E	I	L	P	Q	R	I	Q	Q	
	1251			1260			1269			1278			1287			1296		
TGG	CAG	CAG	AGC	CCT	TGC	ATT	GCT	GAA	GAG	CAC	GGC	AAG	AAG	CTG	CTC	GAA	CGC	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
W	Q	Q	S	P	C	I	A	E	E	H	G	K	K	L	L	E	R	
	1305			1314			1323			1332			1341			1350		
ATT	CGC	CGA	GAG	CAG	CAG	AGT	GCC	CGC	ACC	CGC	CTT	CAG	GAA	ATG	GAA	CGC	CGA	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
I	R	R	E	Q	Q	S	A	R	T	R	L	Q	E	M	E	R	R	
	1359			1368			1377			1386			1395			1404		
TTC	CAT	GAG	CTT	GAG	GCC	ATC	ATT	CTA	CGT	GCC	AAG	CAG	CAG	GCT	GTG	CGC	GAG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
F	H	E	L	E	A	I	I	L	R	A	K	Q	Q	A	V	R	E	
	1413			1422			1431			1440			1449			1458		
GAT	GAG	GAG	AGC	AAC	GAG	GGT	GAC	AGT	GAT	GAC	ACA	GAC	CTG	CAG	ATC	TTC	TGT	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
D	E	E	S	N	E	G	D	S	D	D	T	D	L	Q	I	F	C	
	1467			1476			1485			1494			1503			1512		
GTT	TCC	TGT	GGG	CAC	CCC	ATC	AAC	CCA	CGT	GTT	GCC	TTG	CGC	CAC	ATG	GAG	CGC	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
V	S	C	G	H	P	I	N	P	R	V	A	L	R	H	M	E	R	
	1521			1530			1539			1548			1557			1566		
TGC	TAC	GCC	AAG	TAT	GAG	AGC	CAG	ACG	TCC	TTT	GGG	TCC	ATG	TAC	CCC	ACA	CGC	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
C	Y	A	K	Y	E	S	Q	T	S	F	G	S	M	Y	P	T	R	
	1575			1584			1593			1602			1611			1620		
ATT	GAA	GGG	GCC	ACA	CGA	CTC	TTC	TGT	GAT	GTG	TAT	AAT	CCT	CAG	AGC	AAA	ACA	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
I	E	G	A	T	R	L	F	C	D	V	Y	N	P	Q	S	K	T	
	1629			1638			1647			1656			1665			1674		
TAC	TGT	AAG	CGG	CTC	CAG	GTG	CTG	TGC	CCC	GAG	CAC	TCA	CGG	GAC	CCC	AAA	GTG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Y	C	K	R	L	Q	V	L	C	P	E	H	S	R	D	P	K	V	
	1683			1692			1701			1710			1719			1728		
CCA	GCT	GAC	GAG	GTA	TGC	GGG	TGC	CCC	CTT	GTA	CGT	GAT	GTC	TTT	GAG	CTC	ACG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
P	A	D	E	V	C	G	C	P	L	V	R	D	V	F	E	L	T	
	1737			1746			1755			1764			1773			1782		
GGT	GAC	TTC	TGC	CGC	CTG	CCC	AAG	CGC	CAG	TGC	AAT	CGC	CAT	TAC	TGC	TGG	GAG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
G	D	F	C	R	L	P	K	R	Q	C	N	R	H	Y	C	W	E	

Fig. 2 (continued)

1791			1800			1809			1818			1827			1836		
AAG	CTG	CGG	CGC	GCG	GAA	GTG	GAC	TTG	GAG	CGC	GTG	CGT	GTG	TGG	TAC	AAG	CTG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
K	L	R	R	A	E	V	D	L	E	R	V	R	V	W	Y	K	L
1845			1854			1863			1872			1881			1890		
GAC	GAG	CTG	TTT	GAG	CAG	GAG	CGC	AAT	GTG	CGC	ACA	GCC	ATG	ACA	AAC	CGC	GCG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
D	E	L	F	E	Q	E	R	N	V	R	T	A	M	T	N	R	A
1899			1908			1917			1926			1935			1944		
GGA	TTG	CTG	GCC	CTG	ATG	CTG	CAC	CAG	ACG	ATC	CAG	CAC	GAT	CCC	CTC	ACT	ACC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
G	L	L	A	L	M	L	H	Q	T	I	Q	H	D	P	L	T	T
1953			1962			1971											
GAC	CTG	CGC	TCC	AGT	GCC	GAC	CGC	TGA	3'								
---	---	---	---	---	---	---	---	---									
D	L	R	S	S	A	D	R	*									

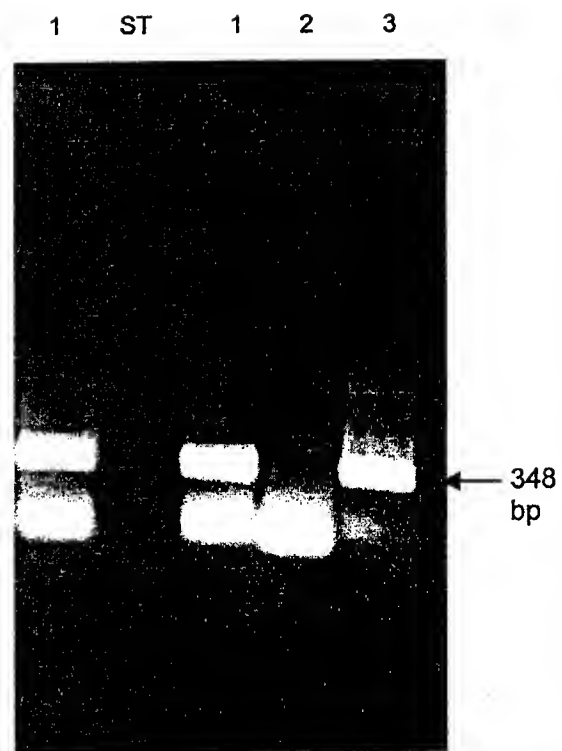
6/10

Fig. 3



7/10

Fig. 4



8/10

Fig. 5



9/10

Fig. 6



10/10

Fig. 7

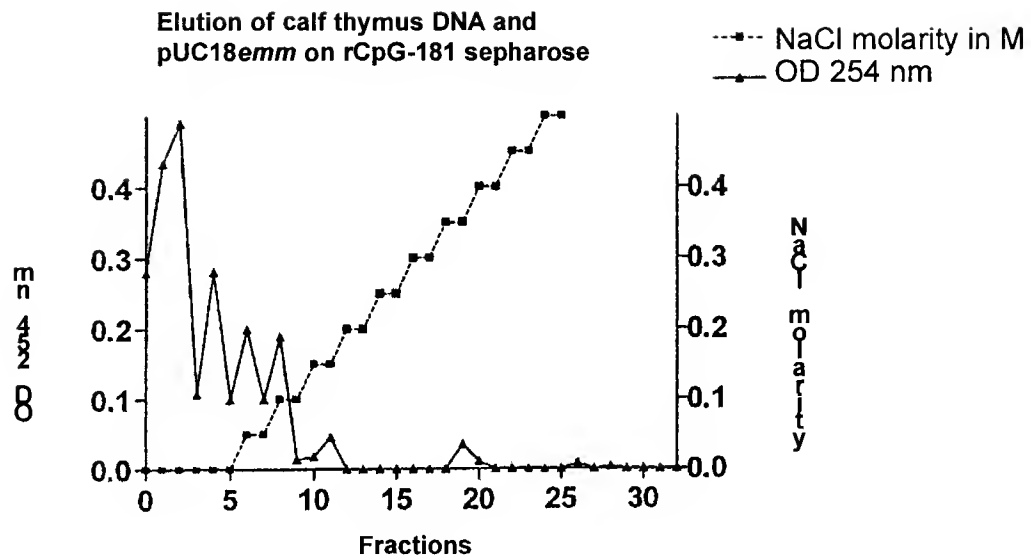


Fig. 8

